Age-moderation of genetic and environmental contributions to cognitive functioning in mid- and late-life for specific cognitive abilities

Age-Moderation of Genetic and Environmental Contributions to Cognitive Functioning in Mid- and Late-Life for Specific Cognitive Abilities

Shandell Pahlen[1]; Nayla R. Hamdi[1]; Anna K. Dahl Aslan[2,3] ; Briana N. Horwitz[4]; Matthew S. Panizzon[5]; Inge Petersen[6]; Catalina Zavala[2,7]; Kaare Christensen[6]; Deborah Finkel[8]; Carol E. Franz[5]; Margaret Gatz[2,9]; Wendy Johnson[10]; William S. Kremen[5]; Robert F. Krueger[1]; Jenae M. Neiderhiser[11]; Chandra A. Reynolds[7]; Nancy L. Pedersen[2,9]; Matt McGue[1,6]

1. Department of Psychology, University of Minnesota, Minneapolis, MN, 55455 USA
2. Department of Medical Epidemiology and Biostatistics, Karolinska Institute, SE-17177 Stockholm, Sweden
3. Jönköping affiliation to School of Health and Welfare, Jönköping University, Jönköping, Sweden
4. Department of Psychology, California State University Fullerton, Fullerton, CA 92834 USA
5. Department of Psychiatry, University of California San Diego, La Jolla, CA 92093 USA
6. The Danish Twin Register and the Danish Aging Research Center, University of Southern Denmark, DK-5000, Odense C, Denmark
7. Department of Psychology, University of California Riverside, Riverside, CA 92521 USA
8. Department of Psychology, Indiana University Southeast, New Albany, IN 47150 USA
9. Department of Psychology, University of Southern California, Los Angeles, CA 90089 USA
10. Department of Psychology and Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Midlothian Scotland
11. Department of Psychology, Penn State University, University Park, PA 16802 USA

Author Contributions: All authors helped to conceptualize the project and design the data analysis through conference calls. All authors read, commented, edited multiple drafts of the paper, and approved the final version. SP, NRH and MM took primary responsibility for undertaking the analyses; SP and MM took primary responsibility for writing the initial draft of the paper. Data were contributed by IP & KC (MADT and LSADT); CEF and WSK (VETSA); RFK (MIDUS); JMN (TOSS); NLP, ADA, CAR, DF, and MG (SATSA, OCTO-Twin and Gender); and MM (MTSADA).

Correspondence: Shandell Pahlen
Department of Psychology
University of California- Riverside
900 University Ave
Riverside, CA 92521
spahl001@ucr.edu

ABSTRACT

Age moderation of genetic and environmental contributions to Digits Forward, Digits Backward, Block Design, Symbol Digit, Vocabulary, and Synonyms was investigated in a sample of 14,534 twins aged 26 to 98 years. The Interplay of Genes and Environment across Multiple Studies (IGEMS) consortium contributed the sample, which represents nine studies from three countries (USA, Denmark, and Sweden). Average test performance was lower in successively older age groups for all tests. Significant age moderation of additive genetic, shared environmental, and non-shared environmental variance components was observed, but the pattern varied by test. The genetic contribution to phenotypic variance across age was smaller for both Digit Span tests, greater for Synonyms, and stable for Block Design and Symbol Digit. The non-shared environmental contribution was greater with age for the Digit Span tests and Block Design, while the shared environmental component was small for all tests, often more so with age. Vocabulary showed similar age-moderation patterns as Synonyms, but these effects were nonsignificant. Findings are discussed in the context of theories of cognitive aging.

Key Words: Aging; Behavior Genetics; Cognitive Ability; Adult Development

INTRODUCTION

Cross-sectional and longitudinal research has consistently found that average cognitive test performance declines in late life (Salthouse, 2009). Nonetheless, there are marked individual differences in the timing and rate of cognitive aging, and late-life cognitive function is relatively etiologically distinct from cognitive function at earlier ages (Wilson et al., 2002). Late-life general cognitive ability (GCA) is also moderately to strongly heritable, with minimal shared environmental contributions (Johnson, McGue, & Deary, 2014). An important but largely unaddressed question concerns whether the magnitudes of genetic and environmental contributions to late-life cognitive ability differ from those at earlier life stages.

A prominent finding from the behavioral genetic literature is that heritability of behavioral phenotypes increases with age. In a meta-analysis of relevant twin studies, Bergen, Gardner, and Kendler (2007) reported that heritability of diverse behavioral phenotypes including anxiety, externalizing psychopathology, social attitudes, and GCA all increased with age. Other research has documented age-related declines in the importance of shared environmental influences for GCA (Haworth et al., 2010). There are, however, several important limitations in this literature. First, most of the research has focused on transitions from childhood to early adulthood; much less is known about the magnitudes of genetic and environmental contributions beyond early adulthood (Tucker-Drob & Briley, 2014). Second, research on cognitive transitions from childhood to early adulthood has focused almost exclusively on GCA rather than specific cognitive abilities, despite evidence of domain specific variation in their developmental trajectories. Third, most of the research has focused on

standardized, rather than raw, components of variance. Greater heritability, a standardized metric, may be a consequence of less raw environmental contribution to variance, greater genetic variance, or both.

The magnitudes of genetic and environmental contributions to late-life cognitive function might differ from those at earlier ages for several reasons. Reduction in evolutionary pressures in late life as compared to other life stages is posited to lead to amplification of stochastic (i.e., random, Finch & Kirkwood, 2000) and epigenetic processes (Fraga et al., 2005). For example, many individual-level factors (i.e., blood pressure, and physical exercise) are associated with late-life cognitive functioning but not with cognitive status at younger ages (Anstey & Christensen, 2000). The cumulative effect of these factors might be reflected by increased environmental contributions to phenotypic variance with age (c.f., Baltes, Reese, & Lipsitt, 1980). Alternatively, changes in the magnitudes of genetic contributions may reflect amplification of existing genetic factors or mechanisms of gene-environment interplay (Reynolds, Finkel, & Zavala, 2013). For instance, genetic factors that protect against environmental influences leading to cognitive decline (e.g., active developmental processes, Scarr & McCartney, 1983) can lead to greater genetic variance in late life. High educational attainment, occupational complexity, and intellectually-stimulating activities may reflect genetically influenced selections that promote cognitive reserve and prevent decline (Bosma et al., 2002).

Behavioral genetic research on cognitive abilities does not always provide consistent evidence for age differences in relative magnitudes of genetic influences. Finkel and Reynolds

(2010) reviewed the behavioral genetic literature on cognitive aging and concluded that heritability of GCA appears to increase through approximately age 60 and declines thereafter. Yet, in a subsequent large cross-sectional study of 2,332 Danish twins age 46 to 96 years, McGue and Christensen (2013) reported that the magnitude of genetic influence on a measure of GCA was stable across age. Unlike the differential patterns observed by independent studies, recent meta-analyses of twin studies have better convergence to the patterns observed. In a recent meta-analysis of twin studies, Reynolds and Finkel (2015) reported that the heritabilities of specific cognitive abilities including verbal, spatial and memory, were largely stable or slightly increasing with age. Similarly, a large-scale meta-analysis of all published twin studies by Polderman, Benyamin, de Leeuw, et al (2015) also found consistent evidence for stable heritability across age groups across cognitive domains of clustered executive functioning and memory abilities. Although these meta-analyses seem to provide a clearer and more consistent pattern of the genetic and environmental contributions to late life, they may be also obscuring differential trajectories for specific cognitive abilities, and indeed losing important informative differences across time.

Limited sample sizes and study and country differences may contribute to the apparent inconsistency of results concerning age moderation of genetic influences. In many cases, heritability of late-life cognitive ability is estimated in samples with a few hundred twin pairs, making it difficult for a single study to distinguish heritability differences across a wide age range reliably. Moreover, studies do not always report parameter estimates for the same biometric model, making it difficult to compare estimates using meta-analytic methods. For

example, the shared environmental contribution is not always reported and some reported heritability estimates are based on models dropping this component.

This study includes 14,534 participants from a twin study consortium to investigate age moderation of genetic and environmental influences on cognitive ability in mid- through late-life. The large sample, broad age range (26 to 98 years), and multiple cognitive abilities included (six tests representing four separate domains of cognitive functioning – short-term/working memory, processing speed, spatial processing, and verbal ability) make this the most comprehensive test to date of the hypothesis that the magnitudes of genetic and environmental influences on cognitive functioning differ in late-life compared to earlier life stages. In addition, the consortium this study is derived from provides a special opportunity to directly assess differential evidence found by independent studies, often from competing independent studies that are included in this consortium group, while simultaneously examining if there are informative differences across time that meta-analytic work may not have been able to observe.

METHOD

Participants

The sample was drawn from nine studies representing three separate countries (Sweden, Denmark, and the United States) from the Interplay of Genes and Environment across Multiple Studies (IGEMS) consortium (Pedersen et al., 2013). No studies had overlapping participants. To be included in our analysis, participants had to have completed at least one of six cognitive tests (described below), and have a Mini-Mental State Examination (MMSE) score of at least 24,

following the typical cutoff for cognitive impairment (Tombaugh & McIntyre, 1992). A total of 1,136 (7.8% of the total number of potential participants) were excluded based on this screen, leaving a sample of 14,534 (50.9% women) individual twins. The sample included 2,341 pairs of monozygotic (MZ) twins, 2,429 pairs of dizygotic-same sex twins (DZ-ss), and 929 pairs of dizygotic-opposite sex twins (DZ-os). The sample also included 3,128 unpaired twins, who were informative with respect to age differences in means and variances and so were included in the analyses. For studies with longitudinal assessments, only data from the first test administration for each participant were used in the cross-sectional analyses reported here. Mean age at that measurement occasion was 61.3 years (Mdn=59.82, *SD* = 13.0). The median was slightly lower than the mean, suggesting a positive skew, although the difference is about a one tenth of a SD. Demographic characteristics for each study, including sample size, gender ratio, age, zygosity and which cognitive tests were administered, are given in Table 1. Figure 1 gives the age distribution of the total sample. Brief descriptions of each of the nine studies, separated by country of origin, are given below. Additional details concerning the methodology for each study can be found in the citations provided.

Table 1. Demographic Characteristics of the Twin Samples

Study	N	% Female	# of Complete Twin Pairs			Mean Age (SD)	Cognitive Test
			MZ	DZ-ss	DZ-os		
Swedish Studies:							
SATSA	788	59.5%	142	221	0	63.6 (8.3)	DF,DB,BD,SD,SYN
GENDER	447	50.8%	0	0	205	74.5 (2.6)	BD,SD,SYN
OCTO-Twin	514	64.6%	97	107	0	83.1 (2.8)	DF,DB,BD,SD,SYN
TOSS	1732	62.9%	380	475	0	44.8 (4.9)	SYN
Danish Studies:							
LSADT	3480	57.2%	325	467	0	76.0 (4.9)	DF,DB,SD
MADT	4280	49.0%	656	589	610	56.4 (6.3)	DF,DB,SD

US Studies:							
VETSA	1230	0.0%	347	263	0	55.4 (2.5)	DF,DB,VOC,SYN
MTSADA	810	60.9%	215	150	0	55.4 (12.6)	BD,SD,VOC
MIDUS	1253	56.0%	179	157	114	54.3 (11.6)	DB
Total	14534	50.9%	2341	2429	929	61.3 (13.0)	DF,DB,BD,SD,VOC,SYN

Note: DF=Digits Forward, DB=Digits Backward, BD=Block Design, SD=Symbol Digit, VOC=Vocabulary, SYN=Synonyms, MZ= Monozygotic, DZ-ss=Dizygotic-same sex, DZ-os=Dizygotic-opposite sex

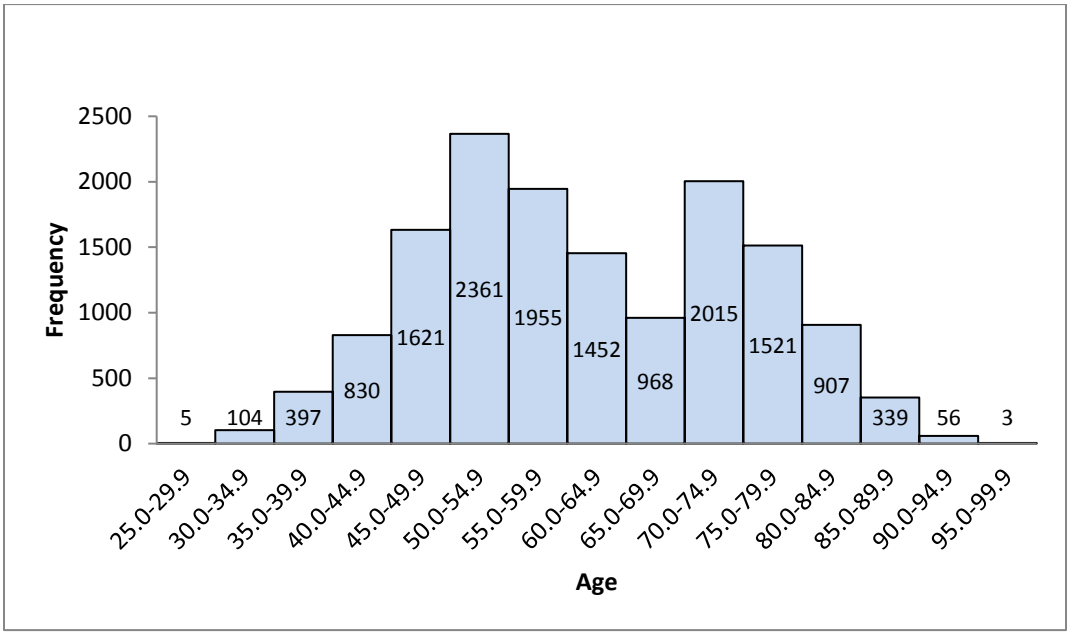

Figure 1: Age distribution of the combined sample of 14,534 individual twins. To qualify for the analyses reported in this paper, a twin needed to complete at least one of the six target cognitive tests and have a MMSE score of at least 24.

Sweden. IGEMS includes four Swedish studies whose samples were all ascertained from records from the Swedish Twin Registry: Swedish Adoption/Twin Study of Aging (SATSA; Pedersen et al., 1991), Aging in Women and Men (GENDER; Gold, Malmberg, McClearn, Pedersen, & Berg, 2002), Origins of Variance in the Oldest-Old (OCTO-Twin; McClearn et al., 1997), and Twin-Offspring Study in Sweden (TOSS; Neiderhiser, Reiss, Lichtenstein, Spotts, & Ganiban, 2007).

Parallel cognitive assessments were used across SATSA, OCTO-Twin and GENDER, and all three studies were longitudinal. The Swedish studies are distinguished by the age range and zygosity represented. SATSA participants include same-sex twins, with a subsample of twins reared apart matched to a subsample of twins reared together by birthdate and county of birth and gender. SATSA in-person testing protocol (IPT) followed a cohort-sequential protocol. Those who had reached age 50 were invited to participate in IPT that began in 1986. At subsequent IPTs, typically conducted at 3-year intervals, SATSA-eligible twins who reached age 50 were invited to participate. Intake cognitive data were collected over four IPT sessions. The age range at initial cognitive testing was 50.0 to 88.0.

GENDER consists of opposite sex twin pairs born between 1906 and 1925. Intake cognitive assessments were completed during a four-year period starting in 1995, when the twins were between 70 and 81 years old.

OCTO-Twin was initiated to investigate same sex twins in very late-life; at their intake assessment twins ranged in age from 79 to 98 years. OCTO-Twin participants completed their intake assessments during a two-year period beginning in 1991.

TOSS was designed to investigate the influences of family relationships within twin families, and so included same sex pairs in which both twins had a teenage child at the time of their intake assessments. Non-twin participants in TOSS, including spouses and offspring, were not included in our analyses. TOSS participants ranged in age from 32 to 59 years.

Denmark. The Danish twin studies include the Longitudinal Study of Aging Danish Twins (LSADT; Christensen, Holm, McGue, Corder, & Vaupel, 1999) and the Middle-Aged Danish Twins Study (MADT; Skytthe et al., 2013). The two Danish studies ascertained twins from the Danish Twin

Registry and administered the same cognitive assessment. LSADT is a cohort-sequential study of same sex twin pairs that began in 1995. Initially, LSADT included twins aged 75 years or older, although as new cohorts were recruited, the age minimum was progressively dropped to 70, age ranges from 70 to 96 years. MADT is a longitudinal study of both same and opposite sex twin pairs born between 1931 and 1952 who were first assessed in 1998, when they ranged in age from 45 to 68 years.

United States. There are three US studies: the Vietnam Era Twin Study of Aging (VETSA; Kremen, Franz, & Lyons, 2013), the Minnesota Twin Study of Adult Development and Aging (MTSADA; Finkel & McGue, 1993), and the twin sample from Midlife Development in the United States (MIDUS; Kendler, Thornton, Gilman, & Kessler, 2000). VETSA is a longitudinal study of a national sample of male twins who served in the military at some time during the Vietnam era (1965-1975) and were recruited through the Vietnam Era Twin Registry. Unlike the other studies, the VETSA sample falls within a restricted age range – all twins were between 51 and 60 years of age at initial assessment, and all twins are male. In MTSADA, twins from same sex twin pairs were ascertained through Minnesota state birth records. MTSADA is a longitudinal study that took place between 1984 and 1994 and included a broad age-range of twins (26 to 87 years at intake), with twins age 60 years and older preferentially recruited. MIDUS is a longitudinal study of a US national sample of middle-age adults that features a twin subsample. The twin sample includes both same and opposite sex twin pairs ascertained using a random-digit dialing procedure and supplemented with additional twin pairs recruited through referrals given by non-twin participants. The MIDUS twin sample ranged in age from 34 to 82 years at the time of cognitive assessment.

Measures and Phenotypic Harmonization

Many of the IGEMS studies administered specific cognitive measures in addition to those analyzed here. We made use of only those measures that could be harmonized across more than one study. This resulted in the inclusion of six different cognitive ability measures: Digits Forward, Digits Backward, Block Design, Symbol Digit, Vocabulary, and Synonyms. These tests covered four cognitive domains: short-term/working memory (Digits Forward and Digits Backward), spatial processing (Block Design), processing speed (Symbol Digit), and verbal comprehension (Synonyms and Vocabulary).

Digits Forward. Five of the studies (SATSA, OCTO-Twin, LSADT, MADT and VETSA) administered a Digits Forward task (total N = 10,206). The two Swedish and two Danish studies used the same procedure, which involved reading a 3- to 9-digit number string and asking the respondent to repeat each string in the correct order as administrated, but had different scoring. In Sweden, the score was the highest string length that the respondent correctly repeated so it could vary from 0 to 9; in Denmark the score was the total number of correctly repeated strings for both trials so it could vary from 0 to 14. The Digits Forward task used in VETSA was based on the Weschler Memory Scale-III (WMS-III, Weschler, 1997) Digit Span subtest. The same basic administration procedure was followed as in the Scandinavian studies, except strings ranged in length from 2 through 11. VETSA used the same scoring procedure as in the Danish studies, so that the total score could range from 0 to 20.

Digits Backward. Six of the studies (SATSA, OCTO-Twin, LSADT, MADT, VETSA and MIDUS) administered a Digits Backward task (total N = 11,442), which required the respondent to

repeat a sequence of numbers in the reverse order as presented. The same Digits Backward task, a string length ranging from 2-8, was used in the two Swedish and two Danish studies, with the same scoring in each study as was used for Digits Forward. The VETSA task was derived from the WMS-III and included two trials at each length from 2 through 10, scored as the total number of correctly repeated strings (varying from 0 to 18). The MIDUS task involved the same procedure, sequence lengths, and scoring as used in the Swedish studies, although it was administered over the telephone rather than in person (Tun & Lachman, 2006).

Block Design. Four of the studies (SATSA, GENDER, OCTO-Twin and MTSADA) administered a Block Design task (total N = 2,320). Block Design is a spatial processing task in which the respondent is asked to reproduce the target two-dimensional geometric shapes using sets of three-dimensional colored blocks. The three Swedish studies used the Kohs Block Design Test (Stone, 1985). There were seven separate trials, each scored on a 0 to 6 basis (scoring depending on the speed, efficiency, and accuracy with which the target was reproduced). In MTSADA, the Block Design subtest from the Weschler Adult Intelligence Scale-Revised (WAIS-R, Weschler, 1981) was administered. This task is very similar to the one used in the other studies, except it involved a total of nine items, each scored for accuracy and speed and summed to give the total score.

Symbol Digit (Digit Symbol). Five studies (SATSA, GENDER, OCTO-Twin, LSADT, and MADT) administered a Symbol Digit task, while a sixth (MTSADA) administered a Digit Symbol task (total N = 8,755). Both tasks measure perceptual speed and accuracy, and so were pooled in analyses reported here. For convenience we designate this combined task as Symbol Digit. The

same task was administered in the three Swedish and two Danish studies. Respondents were given a key containing nine separate two-dimensional geometric symbols that were assigned specific digits between 1 and 9. They were then presented with rows of symbols and asked to state out loud what the correct digit was for each symbol. This Symbol Digit task was administered in two blocks of 50 with 45 seconds allowed for each block. The score was the number of correct trials across the two blocks (varying from 0 to 100). In MTSADA, the Digit Symbol task from the WAIS-R (Weschler, 1981) was administered. Respondents were asked to write down the symbol that corresponded with each target digit. There was one single administration of 90-second duration with a total possible score of 90.

Vocabulary. Two studies (VETSA and MTSADA) administered a Vocabulary test in which respondents were required to generate definitions of words (total N = 2,030). In VETSA, the Vocabulary subtest from the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999) was administered, while MTSADA administered the Vocabulary subtest from the WAIS-R (Weschler, 1981).

Synonyms. Five studies (SATSA, GENDER, OCTO-Twin, TOSS and VETSA) administered a Synonyms test (total N = 4,523). In all cases, the Synonyms test required the respondent to select the word that provided the best synonym to a target word from a set of alternatives. The specific words, alternatives and number of items varied across studies.

Zygosity. The specific methods for zygosity determination varied somewhat from study to study but in most cases involved the use of questionnaires supplemented by DNA analysis to resolve

uncertain cases. In all cases, regardless of the slight differences in methods to determine zygosity, methods have been validated for each study.

Scale Harmonization. Because the numbers and difficulties of items, and scoring and administration procedures varied across studies, it was necessary to place cognitive test scores on a common metric while retaining information about age differences in means and variances. To start, we pooled cognitive tests for studies from the same country that had the same testing procedures. Thus, the two Danish studies were pooled together as were the four Swedish studies. In addition to reducing the number of samples to compare pooling increased the age coverage within each pooled sample. The second step in harmonization involved dividing each study sample into four age groups: below 50, 50 to 59.99, 60 to 69.99, and 70 plus. These age groups were selected to provide adequate sample size in each age group and a single group (50 to 59.99) that existed for every test-study combination. The 50-59.99 group was then used to harmonize the differences in scale across the multiple tests for each of the six abilities. Specifically, for each test-study combination, after removing the main effect of sex, scores were linearly transformed to have a mean of 50.0 and standard deviation of 10.0 (i.e., a T-score metric) in the age 50 to 59.99 age group. In this way, we placed each test on the same scale while retaining information about variance differences across age. Finally, to minimize the impact of outlying observations, we winsorized scores within each age group such that scores greater than $\pm$3 SDs from the age group means were set equal to $\pm$3 SD as appropriate. The frequency of winsorized scores ranged from 0.18% for Block Design to 0.88% for Digits Forward. In a normal distribution we would expect 0.3% of scores to fall outside the $\pm$3 SDs range, so that

there were slight excesses of cases in the tails of the distributions of some of the cognitive tests. To correct for multiple testing, we set the significance threshold at .01, a value slightly larger than the Bonferroni corrected value for 6 independent tests (i.e., .008 = .05/6), because of the average inter-correlation among the cognitive measures were about .30.

Statistical Analysis

We first estimated twin correlations in each of the four age groups, both separately by study as well as pooled across studies. Twin correlations by sex were also examined, but no sex differences were found, thus sex was collapsed by zygosity group. Although we clearly lose information when a quantitative variable such as age is arbitrarily divided into discrete groups, these analyses allowed us to gain a preliminary sense for how twin similarity varied across age group and study. Second, we determined which biometric model (ACE or ADE) best fit the data for each test as seen in Table S2 of the supplement. The ACE model included additive genetic (A), shared environmental (C), and non-shared environmental (E) components of variance; in the ADE model the C component was replaced by a non-additive genetic (D) component. With reared-together twin data, C and D cannot be simultaneously estimated. Under these models, MZ twins share all additive genetic and non-additive genetic effects, while DZ twins share half of the additive effect and one-quarter of the non-additive effect. The shared environment represents the effects of environmental factors that contribute to twin similarity regardless of zygosity and so is shared equally by MZ and DZ twins. The nonshared environment represents

those environmental factors (as well as measurement error) that contribute to twin differences and so is not shared by either MZ or DZ twins.

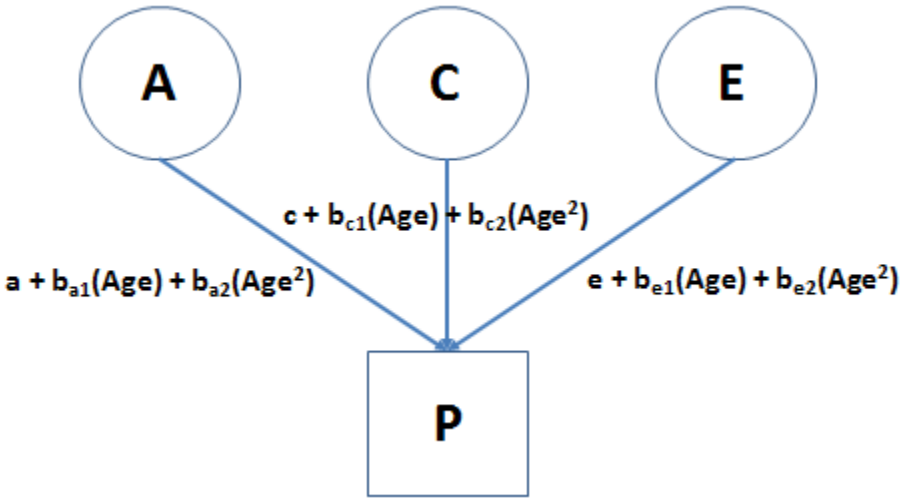

Figure 2: Biometric age-moderation model fit to the twin data. A quantitative phenotype (P) is assumed to be a function of underlying additive genetic (A), shared environmental (C), and non-shared environmental (E) effects. Age moderation is incorporated into the model by allowing the effects of A, C or D (not depicted), and E to be a function of age and squared age.

Lastly, our analyses involved fitting a series of age-moderated biometric models to the pooled twin data for each cognitive test treating age as a continuous variable. The general age-moderation model is depicted in Figure 2 for one member of a twin pair. In this model, we assume that a quantitative phenotype (P) can be expressed as a function of the three factors in the ACE or ADE model, the effect of each potentially moderated by a quadratic function in age.

Because preliminary analysis of DZ twin correlations indicated that correlations for same sex and opposite sex pairs could be pooled without a significant increment in χ^2, no distinction was made between the two types of DZ twins in the age-moderation analyses. For each cognitive outcome, seven models of increasing complexity were fit in Mx (Neale, Boker, Xie, &

Maes, 2004): 1) no-age-moderation, 2) only the total phenotypic variance was moderated by age (i.e., a scalar moderation model implemented by allowing the phenotypic variance to increase as a logistic function of age but constraining the relative contributions of the three biometric components to be invariant across age), 3-5) linear age-moderation of one of the three biometric components only, 6) linear age-moderation model of all three biometric components, and 7) both linear and quadratic age moderation of all three biometric components. Model fit for the biometric models was evaluated using both the χ^2 goodness-of-fit test statistic and the Akaike Information Criteria (AIC = 2lnL+2k, where k is the number of estimated parameters), with models having lower AIC preferred.

RESULTS

Descriptive data

Mean test scores on the T-score scale by age group are reported in Table 2 and depicted in Figure 3. Because of winsorization, the mean and SD in the age 50-59.99 age group deviated slightly from 50 and 10, respectively. Although the age-group effects were statistically significant for all tests, the magnitudes of these effects (as indicated by η^2) were large for Block Design and Symbol Digit (i.e., > 27%), moderate for Synonyms (i.e., 7.7%), and small for Digits Forward, Digits Backward, and Vocabulary (i.e., < 5%). The pattern of age differences was consistent with what has been observed in previous cross-sectional research. That is, Symbol Digit and Block Design showed larger age differences than the verbal tasks (Hoyer, Stawski, Wasylyshyn, & Verhaeghen, 2004) and within the Digit Span tasks, the age effect was greater for Digits Backward than Digits Forward (Bopp & Verhaeghen, 2005).

Table 2. Mean (SD) cognitive test performance as a function of age group pooled across studies

Cognitive Test	Age Group x̄ (SD) N				ANOVA Results F (df) P	Age Group Effect Size
	< 50	50-59	60-69	≥70	Age Group	η^2
Digits Forward	50.8 (10.0) 785	50.0 (9.9) 3498	49.0 (9.6) 1812	48.7 (9.5) 4111	18.5 (3) p < .001	.005
Digits Backward	51.3 (10.1) 1281	50.0 (9.9) 3855	48.6 (9.7) 2065	46.4 (9.5) 4241	132.1 (3) p < .001	.033
Block Design	55.6 (9.8) 221	50.0 (9.9) 428	44.6 (10.4) 554	38.1 (9.6) 1117	287.1 (3) p < .001	.271
Symbol Digit	54.1 (9.1) 1024	49.9 (9.5) 2347	44.7 (10.0) 2104	36.2 (10.0) 3280	1353.2 (3) p < .001	.317
Vocabulary	50.1 (9.7) 259	50.1 (9.8) 1357	47.0 (10.0) 353	45.9 (9.1) 61	12.4 (3) p < .001	.018
Synonyms	48.8 (9.2) 1417	50.1 (9.6) 1836	44.6 (10.6) 293	42.8 (12.2) 977	126.4 (3) p < .001	.077

Note: Scores were transformed within each study to have a mean of 50 and an SD of 10 in the 50-59 age group and then winsorized to ±3SD. ANOVA results take into account the clustering of the twin data.

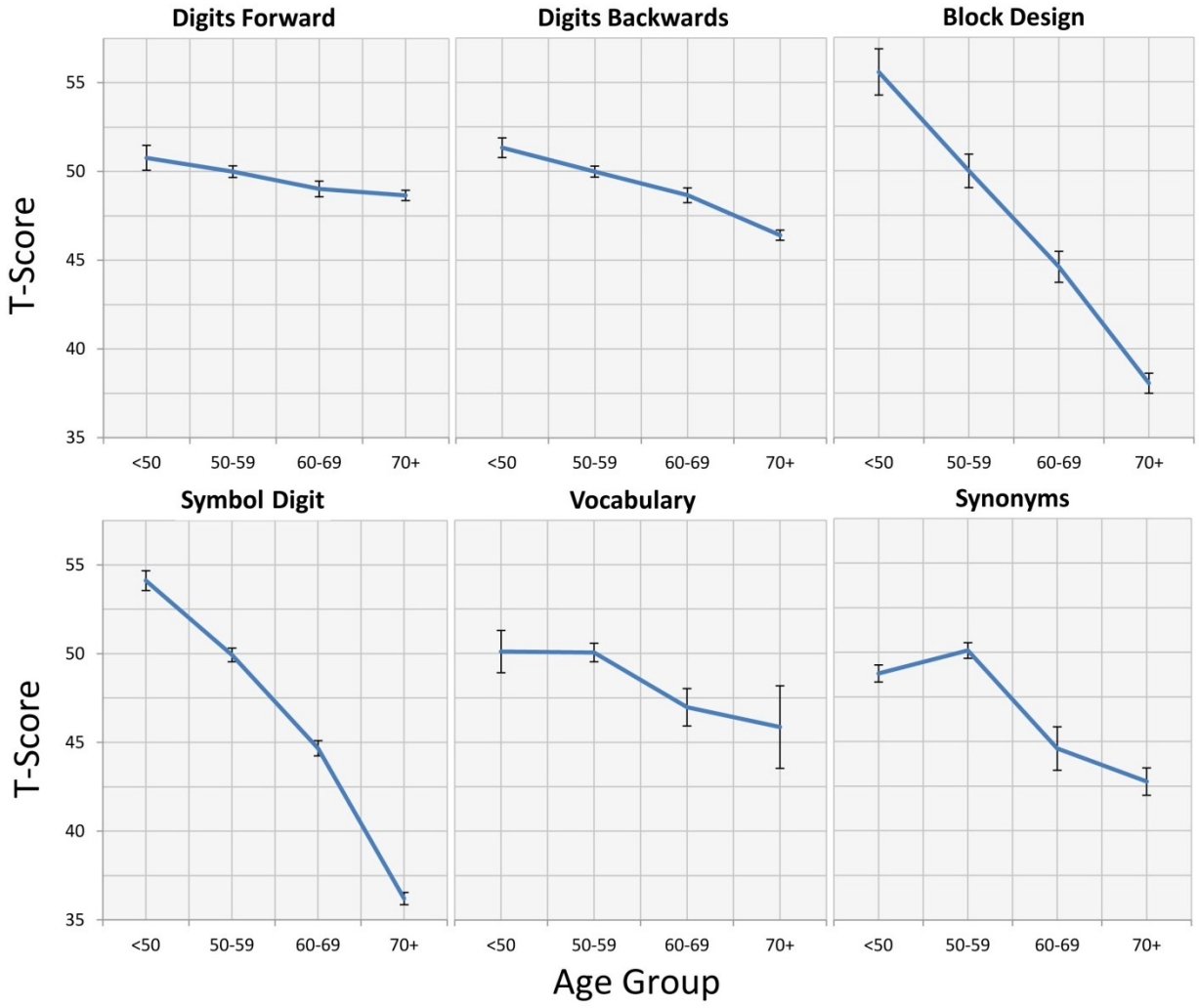

Figure 3: Mean cognitive test score as a function of age group in the pooled sample. Cognitive scores were scaled to a T-score metric such that mean was 50.0 and standard deviation was 10.0 in the 50-59 age group. Error bars give 95% confidence intervals for the means.

Twin Correlations

Table 3 gives the estimated twin correlations by age group pooled across studies. For completeness, all correlations are reported even though the sample sizes in several cells were quite small (i.e., < 25). Also given in the table are the model fit statistics for testing whether the twin correlations varied significantly across studies. Several general trends were noteworthy.

First, there was little evidence for between-study heterogeneity in the twin correlations except for Synonyms, where pooling the correlations resulted in significant increase in the χ^2 test statistic (at p = .001). Although the significant between-study heterogeneity in the twin correlations for Synonyms suggests caution when interpreting results for this scale, for completeness we report modeling results for this phenotype along with the others. Second, decreasing MZ twin correlations with age were observed for several of the cognitive tests (most notably Digits Forward, Digits Backward, and Block Design), suggesting that genetic influences might decline with age for these tests. Alternatively, stable MZ but declining DZ correlations were observed for Vocabulary and Synonyms, suggesting that genetic influences on these tests might increase with age. The purpose of the age-moderation analysis was to formalize these impressions while making maximal use of the available age information.

Table 3. Twin correlations by age-group pooled across studies

		Twin Correlation (95% CI) Number of Pairs		
Cognitive Test	**Age Group**	**MZ**	**DZ-ss**	**DZ-os**
Digits Forward (χ^2 = 15.2 on 10df, p = .13)				
	<50	.36 (.20-.51) 124	.21 (.03-.38) 116	.32 (.13-.47) 102
	50-59	.48 (.42-.54) 695	.30 (.23-.37) 622	.25 (.15-.35) 2913
	60-69	.43 (.33-.52) 285	.25 (.13-.36) 287	.14 (.00-.27) 214
	≥70	.31 (.23-.38) 447	.25 (.18-.32) 593	NA
Digits Backward (χ^2 = 37.1 on 21df, p = .02)				
	<50	.41 (.29-.52) 194	.03 (-.12-.18) 179	.33 (.18-.46) 145
	50-59	.41 (.35-.47) 745	.25 (.17-.32) 664	.19 (.09-.29) 327
	60-69	.41 (.31-.49)	.16 (.05-.26)	.35 (.23-.46)

	323	322	237
≥70	.29 (.21-.37)	.21 (.14-.28)	-.27 (-.73-.40)
	466	608	18
Block Design (χ^2 = 6.5 on 6df, p = .37)			
<50	.71 (.56-.81)	.59 (.33-.77)	NA
	60	34	
50-59	.74 (.63-.82)	.44 (.28-.57)	NA
	85	113	
60-69	.64 (.52-.73)	.32 (.16-.47)	NA
	124	123	
≥70	.54 (.41-.65)	.28 (.11-.42)	.22 (.09-.35)
	124	133	203
Symbol Digit (χ^2 = 21.5 on 16df, p = .16)			
<50	.60 (.50-.68)	.15 (-.01-.30)	.24 (.04-.41)
	190	157	97
50-59	.51 (.44-.58)	.25 (.15-.34)	.27 (.16-.37)
	360	374	281
60-69	.58 (.51-.64)	.35 (.25-.44)	.33 (.20-.45)
	371	333	197
≥70	.56 (.48-.62)	.32 (.23-.40)	.34 (.18-.48)
	335	404	135
Vocabulary (χ^2 = 7.9 on 2df, p = .02)			
<50	.73 (.61-.82)	.74 (.57-.84)	NA
	73	44	
50-59	.63 (.56-.69)	.40 (.30-.49)	NA
	372	289	
60-69	.78 (.69-.84)	.51 (.29-.67)	NA
	98	65	
≥70	.72 (.30-.90)	-.01 (-.54-.52)	NA
	13	14	
Synonyms (χ^2 = 21.7 on 6df, p = .001)			
<50	.68 (.62-.74)	.51 (.43-.58)	NA
	318	383	
50-59	.56 (.49-.62)	.29 (.20-.37)	NA
	463	439	
60-69	.76 (.61-.85)	.30 (.10-.48)	NA
	48	84	
≥70	.73 (.60-.81)	.24 (.06-.41)	.43 (.31-.54)
	83	105	201

NA = Correlation not available; MZ= Monozygotic, DZ-ss=Dizygotic-same sex, DZ-os=Dizygotic-opposite sex. Parenthetical model fit statistics test the fit of pooling twin correlations across the multiple studies.

Age-Moderated Biometric Analysis

Before fitting the age-moderated biometric models, we determined which model, ACE or ADE, provided a better fit for each cognitive test by fitting the alternative three-parameter models to the pooled data. The ACE model had lower AICs for all cognitive tests except Symbol Digit, for which the ADE model was better-fitting (details in the supplementary material). Table 4 gives fit statistics for the age-moderation models based on the best-fitting ACE or ADE model. For each cognitive test we began with a no-age-moderation model and then tested to see whether adding various age-moderation terms improved model fit. The best-fitting model by AIC is highlighted for each test. Several general trends were notable. First, for all cognitive tests except Vocabulary, some form of age-moderation fit better than the no-age-moderation model. That is, when comparing the no-age-moderation model (#1) to the full age-moderation model that included both linear and quadratic age-moderation on the three biometric parameters (#7), the χ^2 test was statistically significant (at $p < .01$) and the AIC for the no-age-moderation model was larger than for the full age-moderation model. Second, only for Symbol Digit was there evidence for quadratic age moderation. That is, the full age-moderation model did not fit better than the linear age-moderation model. Third, except for Vocabulary, a model with age-moderation on the phenotypic variance only (#2) never fit better by AIC than a model with age-moderation on a biometric component. Fourth, for all five cognitive tests for which there was evidence of age moderation, some form of moderation on the genetic component was required and included in the best-fitting model. Because moderation results can be sensitive to extreme scores on the moderator, we repeated all age-moderation analyses with participants' ages

winsorized to 45 and 85 for ages falling below or above these boundaries, respectively. There were very few differences between the models with ages winsorized or preserved. The patterns observed for the variance components had the same trajectories in the two, and the indicated best-fitting models did not differ. Full sample results are reported here.

To determine if the different forms of assessment influenced the results, we replicated the moderation models restricting our analyses to only those studies that had used the exact same test. There were only two tests, Digits Forward and Backward, where this was possible. Overall, the pattern of results for these two tests did not differ markedly when using all available data versus only using data based on the same test. Finally, to assess the power of the age moderation tests, we first derived the median observed non-centrality parameter for each age-moderation model across cognitive tests. For each cognitive test, we then derived the observed power for rejecting the null hypothesis with the effect size set at this median value. Based on this analysis, the observed power was at least 88% for the best-fitting models and ranged from 71% to 100% for the full age-moderation model (i.e., Model #7; details provided in Supplementary material, Table S9).

Graphical displays of the age-moderation results are provided for the raw variance components and phenotypic variances with 95% confidence intervals in Figure 4 and the standardized estimates with 95% confidence intervals in Figure 5. To facilitate comparisons across tests, estimates are reported for the linear moderation on ACE model (#6), except for Symbol Digit where quadratic age moderation for the ADE model (#7) is reported. Several patterns were evident. First, phenotypic variance increased for the two verbal tests, decreased for the two Digit Span tests and Block Design, and was generally stable for Symbol Digit.

Second, Digits Forward and Digits Backward show a near identical pattern of declining genetic variances but stable shared environmental and non-shared environmental variances. As a consequence, the heritability of the two span measures declined from about 35% at age 45 to 20% at age 80. Third, the two verbal tests, Vocabulary and Synonyms, showed similar patterns of increasing genetic variances and declining shared environmental variances across age even though the no-age-moderation model fit best for Vocabulary and the linear moderation on ACE model fit best for Synonyms. The non-shared environmental component was, however, stable with age for Vocabulary but increasing for Synonyms. Heritability estimates for these two tests increased from 30-40% at age 45 to 65-75% at age 85. Third, although heritability of Block Design was relatively stable across age (at approximately 45%), on the raw scale this test showed declining genetic and shared environmental variance and increasing non-shared environmental variance with age. Finally, Symbol Digit, the only test for which modeling analyses supported quadratic age-moderation effects and where the ADE model fit better than the ACE model, showed an increase in additive genetic variance and a decrease in dominance genetic variance up to about age 60 followed by relative stability in these components through age 80. The total genetic variance being the sum of additive and dominance effects, however, was stable across the full age range for Symbol Digit.

Table 4. Fit statistics for age-moderation models

	Model Fit				Improvement in Fit Relative to No-Age-Moderation Model		
Cognitive Test	-2lnl	df	K	AIC	χ^2	Δdf	p
Digits Forward							
1.No Age Moderation	74071.0	10084	13	74097.0			
2. Scalar model	74054.5	10083	14	74082.5	16.5	1	<.001
3. Linear Moderation only on A	74050.1	10083	14	74078.1	20.9	1	<.001
4. Linear Moderation only on C	74054.3	10083	14	74082.3	16.7	1	<.001
5. Linear Moderation only on E	74060.5	10083	14	74088.5	10.5	1	<.001
6. Linear Moderation on ACE	74050.1	10081	16	74082.1	20.9	3	<.001
7. Full ACE Moderation	74047.9	10078	19	74085.9	23.1	6	<.001
Digits Backward							
1.No Age Moderation	83171.5	11297	15	83201.5			
2. Scalar model	83159.8	11296	16	83191.8	11.7	1	<.001
3. Linear Moderation only on A	83155.9	11296	16	83187.9	15.6	1	<.001
4. Linear Moderation only on C	83158.3	11296	16	83190.3	13.2	1	<.001
5. Linear Moderation only on E	83163.6	11296	16	83195.6	7.9	1	0.004
6. Linear Moderation on ACE	83155.9	11294	18	83191.9	15.6	3	<.001
7. Full ACE Moderation	83153.5	11291	21	83195.5	18.0	6	0.006
Block Design							
1.No Age Moderation	17113.9	2306	10	17133.9			
2. Scalar model	17113.2	2305	11	17135.2	0.6	1	0.44
3. Linear Moderation only on A	17106.0	2305	11	17128.0	7.8	1	0.005
4. Linear Moderation only on C	17099.7	2305	11	17121.7	14.2	1	<.001
5. Linear Moderation only on E	17109.8	2305	11	17131.8	4.0	1	0.05
6. Linear Moderation on ACE	17087.4	2303	13	17113.4	26.4	3	<.001
7. Full ACE Moderation	17087.0	2300	16	17119.0	26.8	6	<.001
Symbol Digit							
1.No Age Moderation	64409.3	8711	14	64437.3			
2. Scalar model	64409.0	8710	15	64439.0	0.3	1	0.58
3. Linear Moderation only on A	64408.1	8710	15	64438.1	1.17	1	0.23
4. Linear Moderation only on D	64405.7	8710	15	64435.7	3.61	1	0.06
5. Linear Moderation only on E	64409.2	8710	15	64439.2	0.04	1	0.84
6. Linear Moderation on ADE	64404.3	8708	17	64438.3	5.0	3	0.17
7. Full ADE Moderation	64381.8	8705	20	64421.8	27.4	6	<.001

8. Linear and Quadratic Moderation only on A	64404.7	8709	16	64436.7	4.5	2	0.03
9. Linear and Quadratic Moderation only on D	64386.5	8709	16	64418.5	22.8	2	<.001
10. Linear and Quadratic Moderation only on E	64409.2	8709	16	64441.2	0.1	2	0.98
Vocabulary							
1.No Age Moderation	14613.2	2023	7	14627.2			
2. Scalar model	14613.1	2022	8	14629.1	0.1	1	0.78
3. Linear Moderation only on A	14613.1	2022	8	14629.1	0.1	1	0.73
4. Linear Moderation only on C	14613.1	2022	8	14629.1	0.1	1	0.75
5. Linear Moderation only on E	14613.1	2022	8	14629.1	0.1	1	0.73
6. Linear Moderation on ACE	14610.5	2020	10	14630.5	2.7	3	0.44
7. Full ACE Moderation	14604.0	2017	13	14630.0	9.2	6	0.16
Synonyms							
1.No Age Moderation	33136.5	4503	12	33160.5			
2. Scalar model	33021.2	4502	13	33047.2	115.3	1	<.001
3. Linear Moderation only on A	33030.5	4502	13	33056.5	106.0	1	<.001
4. Linear Moderation only on C	33065.6	4502	13	33091.6	70.9	1	<.001
5. Linear Moderation only on E	33030.6	4502	13	33056.6	105.9	1	<.001
6. Linear Moderation on ACE	33004.4	4500	15	33034.4	132.1	3	<.001
7. Full ACE Moderation	32999.4	4497	18	33035.4	137.1	6	<.001

k is the number of estimated parameters, including means.
AIC is equal to -2lnl+2*k
df = degrees of freedom
A = additive genetic, C = shared environmental, E = non-shared environmental, D = dominance. ACE models fit for all tests except Symbol Digit for which preliminary data suggested non-additivity so an ADE model was fit
1) a no-age-moderation model, 2) a scalar model is only moderation on the total phenotypic variance and all component are kept constant, 3-5) a linear age-moderation model only on one component (i.e., A, C or D, E and no quadratic terms), 6) a linear age-moderation model on the ACE or ADE, 7) a full age-moderation model, including both linear and quadratic terms. Best-fitting model for each cognitive test is highlighted.

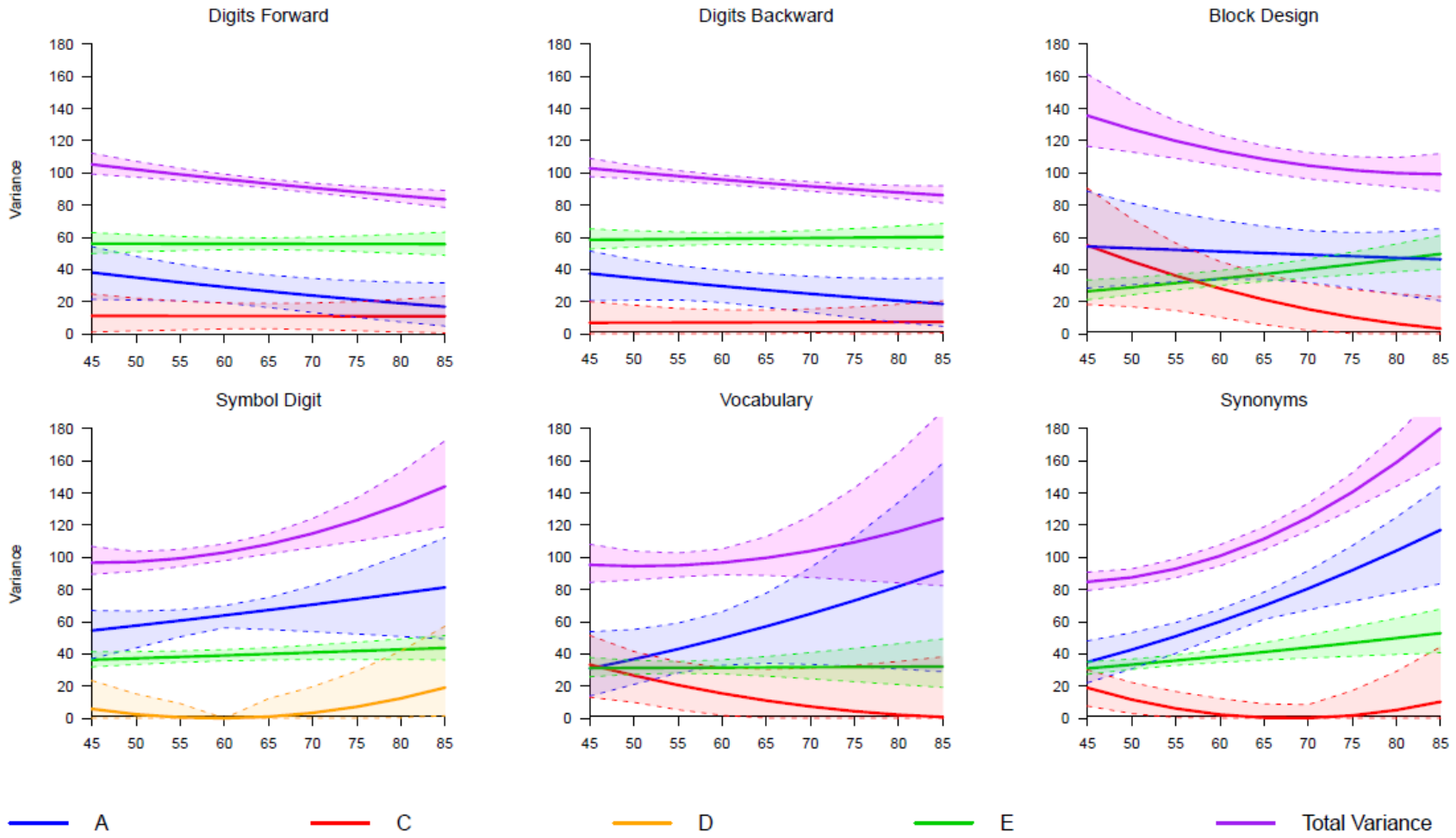

Figure 4: Raw variance component estimates and total phenotypic variance with 95% confidence intervals from the age-moderation model. A = additive genetic component of variance, C = shared environmental variance component, D= dominance variance component and E = non-shared environmental variance component.

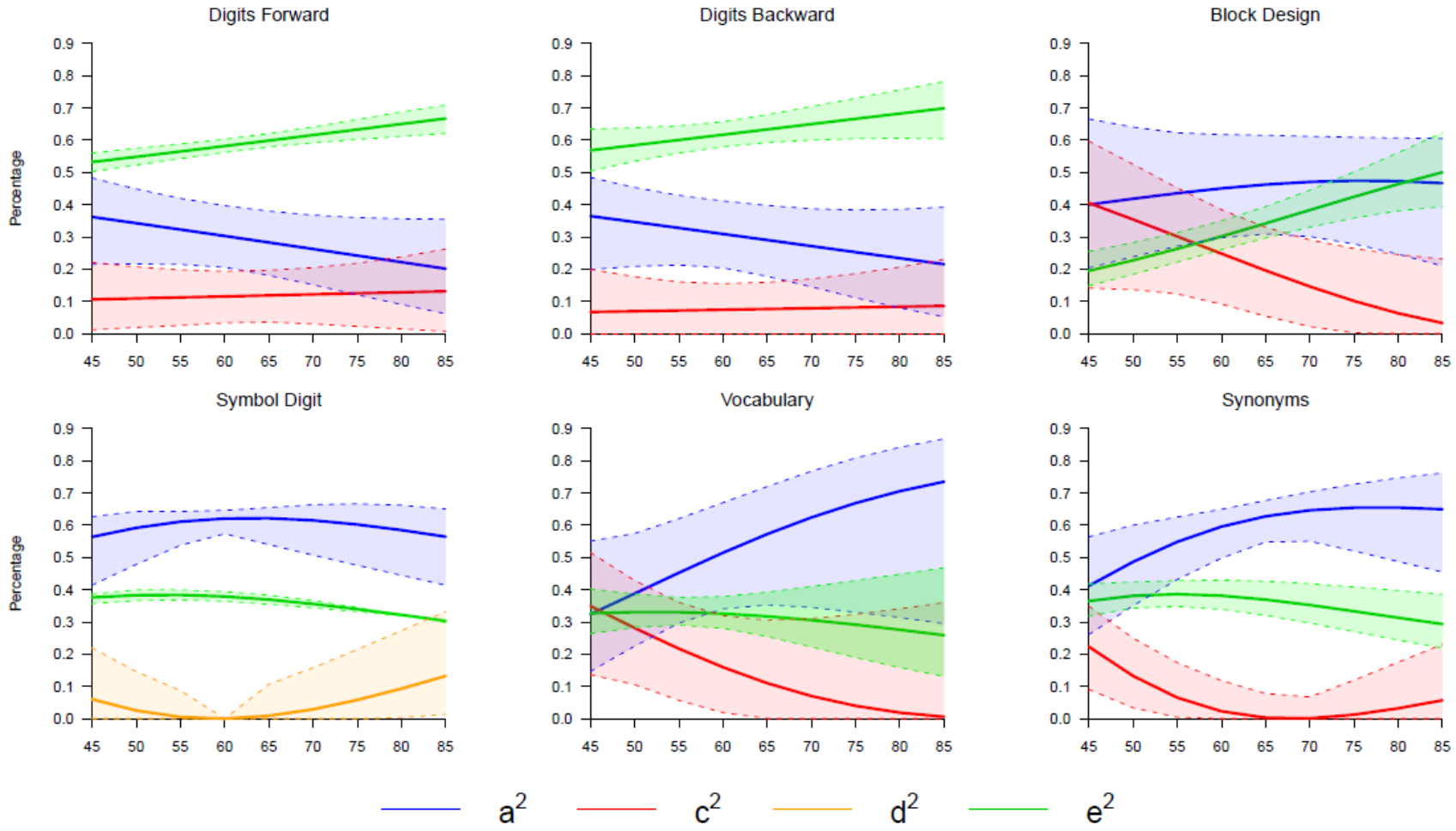

Figure 5: Standardized variance component estimates with 95% confidence intervals from the age-moderation model; a^2 = proportion of phenotypic variance attributed to additive genetic factors, c^2 = proportion of phenotypic variance attributed to shared environmental factors, d^2 = proportion of phenotypic variance attributed to dominance factors, and e^2 = proportion of phenotypic variance attributed to non-shared environmental factors.

DISCUSSION

Cross-sectional analyses of six specific measures of cognitive function in a combined sample of 14,534 twins aged 26 to 98 years revealed age differences in the magnitudes of genetic and environmental contributions to phenotypic variance that varied across test. Before discussing these findings, we acknowledge the limitations of our study. First, pooling cognitive measures across nine twin samples may obscure important between-study differences. Nonetheless, there was limited statistical evidence of between-study heterogeneity in twin similarity, apart from the verbal domain, suggesting that it may not be an important contributor to our results. Second, the twin samples came from three relatively affluent countries, which may limit the generalizability of our findings. Lastly, the six tests we investigated were not all administered in the nine studies nor do they capture the full range of cognitive abilities implicated in aging. Consequently, we could not harmonize a measure of GCA; however, our results are still informative about GCA. As a single measure, GCA can, of course, have only one best-fitting model with respect to age moderation from mid- to late life. Therefore, our results indicate that examination of age-moderation effects on genetic and environmental variance in GCA must obscure differing directional change for the various specific cognitive abilities that underlie GCA. GCA typically represents a composite of multiple specific cognitive measures. Our results suggest that these specific measures show different patterns of age moderation of the underlying biometric components of variance. Yet, as a single measure, GCA can only show a single pattern of age moderation, necessarily misrepresenting the diverse patterns of the cognitive abilities that go into its computation.

Despite the above mentioned limitations, our findings provide a comprehensive and informative picture of how genetic influences on cognitive abilities vary across adulthood. Overall, the heritability estimates we report as seen in Table 3 are consistent with past research on specific cognitive abilities (i.e., 23-62%, Reynolds & Finkel, 2015). Research with younger samples has consistently found that the heritability of GCA increases through early adulthood (Bergen, et al., 2007; Haworth, et al., 2010), but it is not known whether the same is true for specific cognitive abilities. We sought to address if there is any difference in heritability of cognitive abilities across age in late life. There are a limited number of twin studies of cognitive function in mid- to late-adulthood, and the studies that do exist have reported increasing, decreasing, as well as stable patterns of heritability with age (Finkel & Reynolds, 2010). This heterogeneity may be the result of different cognitive abilities having different patterns of heritability with age, a possibility that is supported by our findings. Importantly, a pattern of changing heritability could arise because the underlying biometric components are not all changing at the same rate. For example, increasing heritability can be the result of an increasing genetic component but stable environmental components, or alternatively, decreasing environmental components but a stable genetic component. To resolve these possibilities, it is important to investigate age differences in the underlying biometric components, as we did in this study.

The two verbal tasks, Vocabulary and Synonyms, showed age patterning consistent with what has been observed at earlier developmental stages: greater A (additive genetic variance) at higher ages combined with less in C (shared environmental variance). Although we interpret this common pattern here, it is important to recognize that we did not find statistical evidence

for significant age moderation for Vocabulary. At younger developmental stages, this pattern of increasing A with decreasing C has been hypothesized to be due to a shift from passive gene-shared environment to active gene-non-shared environment correlational processes (Scarr & McCartney, 1983). That is, as individuals achieve greater independence in early adulthood, they increasingly seek out environments and experiences that are consistent with and reinforce their underlying genetic dispositions. This active matching of environment to genotype would result in amplification of genetic variance and consequent increases in genetic variance (Briley & Tucker-Drob, 2013). By extension, many estimates of genetic influences in cross-sectional studies reported to date are biased. The results reported here are consistent with active genotype-environment correlational processes continuing to be important throughout mid- and late-life for verbal-related abilities. The finding that intellectual engagement in midlife was significantly heritable (McGue, Skytthe, & Christensen, 2014) suggests that intellectual engagement might be a useful target in future investigations of the role of active gene-environment correlational processes in late-life cognitive functioning. Why this process would apply to only two of the six cognitive tests we investigated is, however, unclear. One possibility is that word knowledge may be amenable to active gene-environment correlation because it is amenable to routine practice (e.g., through reading books), while the other cognitive tasks we assessed are typically not required in everyday life. Regardless, resolving the basis for greater genetic variance with age observed with the verbal tests will ultimately require analysis of longitudinal data.

Although the age moderation models for vocabulary seem to provide an interesting pattern of genetic and environmental contributions, we want to be cautious in interpreting this

result to discuss why this was the only one of six measures in which age moderation was non-significant. Low power may have contributed to our failure to find evidence for age moderation. Vocabulary had the smallest sample, N=968, of all cognitive tests examined. We did find significant age moderation effects for Block Design, which had only a slightly larger sample size (N=996), but there were only 27 people in the >70 age group for Vocabulary compared with 257 for Block Design. Notably, 68% of the sample for Vocabulary fell in the 50-59 age range, suggesting that the combination of relatively small sample size and low age variability limited the power for Vocabulary. Consistent with this interpretation, among all tests the observed statistical power under the full ACE moderation model (#7) was lowest for Vocabulary (71%, Supplementary Table S9).

A common interpretation of the lifespan developmental perspective leads to the expectation that the E (non-shared environmental variance) contribution to cognitive function will increase with age (Baltes, et al., 1980). We found evidence for notable age increases in E for only two cognitive tests: Block Design and Synonyms. This increase in E is further strengthened when A is either stable or decreasing, a pattern observed for Block Design. Since E is confounded with measurement error, however, an increase in E could be due to declining reliability rather than increasing importance of unique experiential factors. We indirectly evaluated the possibility of differential reliability by assessing the inter-test correlations in all of the four age groups for five test pairings (Table S1 in the supplemental materials). The resulting average phenotypic correlations ranged from .34 in the less than 50 group to .42 in the 70 and older group, suggesting little differential attenuation caused by measurement error. This observation was also supported by the stable split-half correlations reported by the WAIS-R for

Block Design in age groups covering 16 to 74 (Weschler, 1981). Our results thus suggested that Block Design and Synonyms might be usefully targeted to identify the unique experiential factors that contribute to individual differences in late-life cognitive functioning. Alternatively, greater E with age may signal greater differences in genetic sensitivity to environmental exposures (Reynolds, et al., 2013), as in standard biometric models of reared-together twin data gene by non-shared environmental interaction effects are included in the estimate of E.

Although we observed greater E with age for Block Design and Synonyms, A for these tests was either slightly lower (Block Design) or notably greater with age (Synonyms). When these patterns were extended to the standardized model (as seen in Figure 5), heritability was observed to be stable and greater with age for Block Design and Synonyms, respectively. We only observed smaller heritability with age for the two Digit Span tasks, and in both cases this owed to smaller A rather than to greater E. This pattern was unexpected but does mirror a similar pattern found for GCA after age 60 by Finkel and Reynolds (2010). The decrease in the genetic component of variance for the span measures could arise if task performance strategies change with age, possibly due to perceptions about memory function. Longitudinal studies are needed to determine whether the decline in genetic influence for the span measures is due to different genetic factors influencing span performance at different ages or the same genetic factors having different magnitudes of effect at different ages.

The genetic contributions to Block Design and Symbol Digit were relatively stable with age, which is in contrast to the lower heritability of Digit Span measures, or the greater heritability of the verbal measures with age. For Symbol Digit, the only test for which the ADE model fit best, E was stable with age, but A increased through age 70 while the dominance

variance component (D) declined after age 60. As a consequence, the total genetic contribution (i.e., additive plus dominance) on both the raw and standardized scale was largely stable for this test. For Block Design the heritability increased slightly with age, despite slightly less A with age but substantially greater E. This seemingly anomalous pattern was the consequence of sharply less C with age for Block Design. Less C with age was observed with all three tests, Block Design, Vocabulary, and Synonyms, for which there was evidence of moderate shared environmental influences at the youngest ages in our sample. This pattern is consistent with the general finding in behavioral genetics that the shared environment becomes diminishingly important the further one is temporally removed from the rearing home (Lichtenstein, Pedersen, & McClearn, 1992).

In summary, we found evidence of age moderation of the biometric components of variance that differed across cognitive tasks. Considering proportions of variance, we observed increasing heritability for Vocabulary and Synonyms, decreasing heritability for Digits Forward and Digits Backward, and stable heritability for Block Design and Symbol Digit. We found support for the prediction from lifespan developmental theory that unique experiences become increasingly important with age for Block Design and Synonyms. Alternatively, for the two verbal tasks, Vocabulary and Synonyms, we found evidence consistent with the increasing importance of active gene-environment correlational processes with age.

This study provides a preliminary step on moving closer to filling the gaps within the behavior genetic literature on cognitive abilities trajectories in late life but more work is still needed. In the future, we plan to extend our analyses to longitudinal data to explore intra-individual variability across age and time, an opportunity available because most of the IGEMS

studies are longitudinal. Relatedly, potential birth cohort effects not examined here should be explored in subsequent work. In addition, to better understand the mechanisms influencing the differential developmental pathways for specific cognitive measures, possible gene-environment interactions should be explored. In conclusion, this study aids in trying to bring convergence on a topic within the literature often plagued by discrepant findings. This study also highlights the importance of the investigating specific cognitive traits across age since genetic and environmental influences varied by measure. Lastly, this study shows that large scale harmonization projects are possible and can provide nuanced and informative findings potentially obscured by standard meta-analytic approaches or unavailable to smaller individual studies.

REFERENCES

Anstey, K., & Christensen, H. (2000). Education, activity, health, blood pressure and apolipoprotein E as predictors of cognitive change in old age: A review. *Gerontology, 46*(3), 163-177. doi:10.1159/000022153

Baltes, P. B., Reese, H. W., & Lipsitt, L. P. (1980). Life-span developmental psychology. *Annual Review of Psychology, 31*, 65-110.

Bergen, S. E., Gardner, C. O., & Kendler, K. S. (2007). Age-related changes in heritability of behavioral phenotypes over adolescence and young adulthood: A meta-analysis. *Twin Research and Human Genetics, 10*(3), 423-433.

Bopp, K. L., & Verhaeghen, P. (2005). Aging and verbal memory span: A meta-analysis. *Journals of Gerontology Series B-Psychological Sciences and Social Sciences, 60*(5), P223-P233.

Bosma, H., van Boxtel, M. P. J., Ponds, R., Jelicic, M., Houx, P., Metsemakers, J., & Jolles, J. (2002). Engaged lifestyle and cognitive function in middle and old-aged, non-demented persons: a reciprocal association? *Zeitschrift Fur Gerontologie Und Geriatrie, 35*(6), 575-581. doi:10.1007/s00391-002-0080-y

Briley, D. A., & Tucker-Drob, E. M. (2013). Explaining the Increasing Heritability of Cognitive Ability Across Development A Meta-Analysis of Longitudinal Twin and Adoption Studies. *Psychological Science, 24*(9), 1704-1713. doi:10.1177/0956797613478618

Christensen, K., Holm, N. V., McGue, M., Corder, L., & Vaupel, J. W. (1999). A Danish population-based twin study on general health in the elderly. *Journal of Aging and Health, 11*, 49-64.

Finch, C. E., & Kirkwood, T. B. L. (2000). *Chance, development, and aging*. New York: Oxford University Press.

Finkel, D., & McGue, M. (1993). The origins of individual differences in memory among the elderly: A behavior genetic. *Psychology and Aging, 8*, 527-537.

Finkel, D., & Reynolds, C. A. (2010). Behavioral Genetic Investigations of Cognitive Aging. In Y. K. Kim (Ed.), *Handbook of Behavior Genetics* (pp. 101-112). New York: Springer.

Fraga, M. F., Ballestar, E., Paz, M. F., Ropero, S., Setien, F., Ballestart, M. L., . . . Esteller, M. (2005). Epigenetic differences arise during the lifetime of monozygotic twins. *Proceedings of the National Academy of Sciences of the United States of America, 102*(30), 10604-10609. doi:10.1073/pnas.0500398102

Gold, C. H., Malmberg, B., McClearn, G. E., Pedersen, N. L., & Berg, S. (2002). Gender and health: A study of older unlike-sex twins. *Journals of Gerontology Series B-Psychological Sciences and Social Sciences, 57*(3), S168-S176.

Haworth, C. M. A., Wright, M. J., Luciano, M., Martin, N. G., de Geus, E. J. C., van Beijsterveldt, C. E. M., . . . Plomin, R. (2010). The heritability of general cognitive ability increases linearly from childhood to young adulthood. *Molecular Psychiatry, 15*(11), 1112-1120. doi:10.1038/mp.2009.55

Hoyer, W. J., Stawski, R. S., Wasylyshyn, C., & Verhaeghen, P. (2004). Adult age and digit symbol substitution performance: A meta-analysis. *Psychology and Aging, 19*(1), 211-214. doi:10.1037/0882-7974.19.1.211

Johnson, W., McGue, M., & Deary, I. J. (2014). Normative cognitive aging. In D. Finkel & C. A. Reynolds (Eds.), *Behavior genetics of cognition across the lifespan* (pp. 135-167). New York: Springer.

Kendler, K. S., Thornton, L. M., Gilman, S. E., & Kessler, R. C. (2000). Sexual orientation in a US national sample of twin and nontwin sibling pairs. *American Journal of Psychiatry, 157*(11), 1843-1846. doi:10.1176/appi.ajp.157.11.1843

Kremen, W. S., Franz, C. E., & Lyons, M. J. (2013). VETSA: The Vietnam Era Twin Study of Aging. *Twin Research and Human Genetics, 16*(1), 399-402. doi:10.1017/thg.2012.86

Lichtenstein, P., Pedersen, N. L., & McClearn, G. E. (1992). The Origins of Individual Differences in Occupational Status and Educational Level A Study of Twins Reared Apart and Together. *Acta Sociologica*, *35*(1), 13-31.

McClearn, G. E., Johansson, B., Berg, S., Pedersen, N. L., Ahern, F., Petrill, S. A., & Plomin, R. (1997). Substantial genetic inlfuence on cognitive abilities in twins 80 or more years old. *Science, 276*, 1560-1563.

McGue, M., & Christensen, K. (2013). Growing old but not growing apart: Twin similarity in the latter half of the lifespan. *Behavior genetics*, *43*(1), 1-12.

McGue, M., Skytthe, A., & Christensen, K. (2014). The nature of behavioural correlates of healthy ageing: a twin study of lifestyle in mid to late life. *International Journal of Epidemiology, 43*(3), 775-782. doi:10.1093/ije/dyt210

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (2004). *Mx: Statistical modeling* (6th ed.). Box 126 MCV, Richmond VA 23298: Department of Psychiatry.

Neiderhiser, J. A., Reiss, D., Lichtenstein, P., Spotts, E. L., & Ganiban, J. (2007). Father-adolescent relationships and the role of genotype-environment correlation. *Journal of Family Psychology, 21*(4), 560-571. doi:10.1037/0893-3200.21.4.560

Pedersen, N. L., Christensen, K., Dahl, A. K., Finkel, D., Franz, C. E., Gatz, M., . . . Reynolds, C. A. (2013). IGEMS: The Consortium on Interplay of Genes and Environment Across Multiple Studies. *Twin Research and Human Genetics, 16*(1), 481-489. doi:10.1017/thg.2012.110

Pedersen, N. L., McClearn, G. E., Plomin, R., Nesselroade, J. R., Berg, S., & de Faire, U. (1991). The Swedish Adoption/Twin Study of Aging: An update. *Acta Geneticae Medicae et Gemellologiae, 40*, 7-20.

Polderman, T. J., Benyamin, B., de Leeuw, C. A., Sullivan, P. F., van Bochoven, A., Visscher, P. M., & Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature genetics, 47*(7),702-709.

Reynolds, C. A., & Finkel, D. (2015). A Meta-analysis of Heritability of Cognitive Aging: Minding the “Missing Heritability” Gap. *Neuropsychology review*, *25*(1), 97-112.

Reynolds, C. A., Finkel, D., & Zavala, C. (2013). Gene by environment interplay in cognitive aging. In D. Finkel & C. A. Reynolds (Eds.), *Behavior genetics of cognition across the lifespan* (pp. 169-199). New York: Springer.

Salthouse, T. A. (2009). When does age-related cognitive decline begin? *Neurobiology of Aging, 30*(4), 507-514. doi:10.1016/j.neurobiolaging.2008.09.023

Scarr, S., & McCartney, K. (1983). How people make their own environments: A theory of genotype-environment effects. *Child Development, 54*(2), 424-435.

Skytthe, A., Christiansen, L., Kyvik, K. O., Bodker, F. L., Hvidberg, L., Petersen, I., . . . Christensen, K. (2013). The Danish Twin Registry: Linking Surveys, National Registers, and Biological Information. *Twin Research and Human Genetics, 16*(1), 104-111. doi:10.1017/thg.2012.77

Stone, M. (1985). Kohs Block Design Test. In D. J. Keyser & R. C. Sweetland (Eds.), *Test Critiques II*. Kansas City: Test Corporation of America.

Tombaugh, T. N., & McIntyre, N. J. (1992). The Mini-Mental State Examination: A comprehesnive review. *Journal of the American Geriatrics Society, 40*(9), 922-935.

Tucker-Drob, E. M., & Briley, D. A. (2014). Continuity of Genetic and Environmental Influences on Cognition Across the Life Span: A Meta-Analysis of Longitudinal Twin and Adoption Studies. *Psychological Bulletin, 140*(4), 949-979. doi:10.1037/a0035893

Tun, P. A., & Lachman, M. E. (2006). Telephone assessment of cognitive function in adulthood: the Brief Test of Adult Cognition by Telephone. [Article]. *Age and Ageing, 35*(6), 629-632. doi: 10.1093/ageing/afl095

Weschler, D. (1981). *Manual for the Weschler Adult Intelligence Scale - Revised*. New York: Psychological Corporation.

Weschler, D. (1997). *Wechsler Memory Scale—Third Edition (WMS–III) administration and scoring manual*. San Antonio, TX: The Psychological Corporation.

Weschler, D. (1999). *Manual for the Wechsler Abbreviated Intelligence Scale (WASI)*. San Antonio*:* The Psychological Corporation.

Wilson, R. S., Beckett, L. A., Barnes, L. L., Schneider, J. A., Bach, J., Evans, D. A., & Bennett, D. A. (2002). Individual differences in rates of change in cognitive abilities of older persons. *Psychology and Aging, 17*, 179-193.

Supplementary Material

Age-Moderation of Genetic and Environmental Contributions to Cognitive Functioning in Mid- and Late-Life for Specific Cognitive Abilities

Contents:

I. Inter-test Correlations
   a. Table S1: Inter-test correlations for each cognitive test by age group
II. Univariate Biometric Analysis
   a. Table S2: Fit statistics for univariate biometric models
III. Supplemental Methods
   a. Table S3: Digits Forward test version by study across age groups
   b. Table S4: Digits Backward test version by study across age groups
   c. Table S5: Block Design test version by study across age groups
   d. Table S6: Symbol Digit test version by study across age groups
   e. Table S7: Vocabulary test version by study across age groups
   f. Table S8: Synonyms test version by study across age groups
IV. Calculating Observed Power
   a. Table S9. Fit statistics for age-moderation models

I. Inter-test Correlations

Table 1 shows the inter-test correlations across each cognitive test and age group. All tasks were compared to one another, but only one of the six tests, Synonyms, did not have full coverage across the age groups. Thus, correlations are only shown for tests representing every age group. The average across the age groups and tests were weighted based on the sample available. There are a couple things to note from this table. First the correlations seen between each test and in each age group tend to be moderate. Second, for each cognitive test correlation, as age increases within the age groups the correlations either remain stable or slightly increase, except for Digits Backward and Forward which slightly decrease. Third, the average inter-test correlation across the age groups is stable and shows little evidence for attenuation.

Table S1: Inter-test correlations for each cognitive test by age group

Cognitive Test	<50 N	50-59 N	60-69 N	70+ N	$\bar{x}$ across Age Group
Digits Forward & Backward	.54 785	.50 3493	.47 1812	.44 4098	.48 10188
Digit Forward & Symbol Digit	.34 767	.23 2213	.29 1751	.25 2822	.27 7553
Digit Backward & Symbol Digit	.36 767	.34 2213	.36 1751	.38 2820	.36 7551
Symbol Digit & Block Design	.32 220	.51 427	.53 551	.62 937	.50 2135
	.12	.14	.40	.42	.27

Symbol Digit & Vocabulary	257	132	351	61	801
Age Group x̄	.34 2796	.34 8478	.41 6216	.42 10738	

II. Univariate Biometric Analysis

The fit statistics for the univariate biometric analysis are shown in Table 2. The additive model (ACE) and the non-additive model (ADE) were compared to the base model to determine which model was best fitting. The base model in this table refers to the unconstrained model, which allows each tests' component estimates to vary by study and age group. The parameter estimates are then pooled across the age groups for each base model. Since this comparison is between two base models with the same number of degrees of freedom, the best fitting model was the model with the lower -2lnl. In all the tests, except for Symbol Digit, the ACE model had the lower -2lnl and was the better fitting model. Conversely, a non-additive model was found to be the better fitting model for only Symbol Digit.

Table S2: Fit statistics for univariate biometric models

	-2lnl	**df**
Digit Forwards		
1. ACE	73991.82	10049
2. ADE	74005.46	10049
Digit Backwards		
1. ACE	83015.45	11240
2. ADE	83017.35	11240
Block Design		
1. ACE	16710.16	2261
2. ADE	16715.11	2261

Symbol Digit		
1. ACE	63503.49	8661
2. ADE	63503.32	8661
Vocabulary		
1. ACE	14562.09	2005
2. ADE	14578.72	2005
Synonyms		
1. ACE	32897.95	4479
2. ADE	32906.91	4479

III. Supplemental Methods

Digits Forward. Five of the studies (SATSA, OCTO-Twin, LSADT, MADT and VETSA) administered a Digits Forward task (total N = 10,208). The two Swedish and two Danish studies used the same procedure, which involved reading a 3- to 9-digit number string and asking the respondent to repeat each string in turn in the correct order. Testing began with the shortest string, consisted of two different strings at each length, and proceeded until the respondent could not correctly repeat either string at a given length. Although the assessment procedure was the same in Sweden and Denmark, the task was scored differently in the two countries. In Sweden, the score was the highest string length that the respondent correctly repeated so it could vary from 0 to 9; in Denmark the score was the number of correctly repeated strings so it could vary from 0 to 14. The Digits Forward task used in VETSA was based on the Weschler Memory Scale-III (WMS-III, Weschler, 1997) Digit Span subtest. The same basic administration procedure was followed as in the Scandinavian studies, except strings ranged in length from 2 through 11. VETSA used the same scoring procedure the as the Danish studies, so that the total score could range from 0 to 20.

Table S3: Digits Forward test version by study across age groups

	Age Groups

Studies	Test	Birth Year Range	Age Range	<50		50-59		60-69		70+	
				N	%	N	%	N	%	N	%
SWEDISH	Dureman-SSlde Battery subtest	1893-1948	50-98	0	0%	145	8%	135	15%	278	26%
DANISH	Dureman-SSlde Battery subtest	1899-1952	45-96	342	66%	859	49%	652	74%	761	70%
VETSA	Wechsler Abbreviated Scale of Intelligence	1944-1955	51-59	0	0%	602	35%	0	0%	0	0%
TOTAL				521	1	1740	1	884	1	1088	100%

Digits Backward. Six of the studies (SATSA, OCTO-Twin, LSADT, MADT, VETSA and MIDUS) administered a Digits Backward task (total N = 11,439), which required the respondent to repeat a sequence of numbers in the reverse order of that presented. The same Digits Backward task was used in the two Swedish and two Danish studies, with two strings read at each length from 2 through 8 until a respondent could not repeat either string. As with Digits Forward, the score in the Swedish studies was based on the longest string that could be repeated in reverse order (varying from 0 to 8) and was the sum of correct responses in the Danish studies (varying from 0 to 14). The VETSA task was derived from the WMS-III and included two trials at each length from 2 through 10, scored as the total number of correctly repeated strings (varying from 0 to 18). The MIDUS task was based on the Wechsler Adult Intelligence Scale-III and included as part of the Brief Test of Adult Cognition by Telephone (Tun & Lachman, 2006). This task involved the same procedure, sequence lengths, and scoring as used in the Swedish studies, although it differed in that it was administered over the telephone rather than in person.

Table S4: Digits Backward test version by study across age groups

	Birth Year	Age	Age Groups <50	50-59	60-69	70+

Studies	Test	Range	Range	N	%	N	%	N	%	N	%
SWEDISH	Dureman-SSlde Battery subtest	1893-1948	50-98	0	0%	145	8%	135	15%	278	26%
DANISH	Dureman-SSlde Battery subtest	1899-1952	45-96	342	66%	859	49%	652	74%	761	70%
VETSA	Wechsler Abbreviated Scale of Intelligence	1944-1955	51-59	0	0%	602	35%	0	0%	0	0%
MIDUS	Brief Test of Adult Cognition by Telephone (BTACT)	1921-1970	34-82	179	34%	134	8%	97	11%	49	5%
TOTAL				521	1	1740	1	884	1	1088	100%

Block Design. Four of the studies (SATSA, GENDER, OCTO and MTSADA) administered a Block Design task (total N = 2,303). Block Design is a spatial processing task in which the respondent is asked to reproduce target two-dimensional geometric shapes using sets of three-dimensional colored blocks. The three Swedish studies used the Kohs Block Design Test (Stone, 1985). There were seven separate trials, each scored on a 0 to 6 basis (scoring depending on the speed and efficiency with which the target was reproduced), so that total scores ranged from 0 to 42. In MTSADA, the Block Design subtest from the Weschler Adult Intelligence Scale-Revised (WAIS-R, Weschler, 1981) was administered. This task is very similar to the one used in the other studies, except it involved a total of nine items, each scored for accuracy and speed and summed to give the total score.

Table S5: Block Design test version by study across age groups

				Age Groups							
		Birth Year	Age	<50		50-59		60-69		70+	
Studies	Test	Range	Range	N	%	N	%	N	%	N	%
SWEDISH	Koh's Block Design	1900-1948	50-93	0	0%	145	73%	131	53%	453	98%
MTSADA	Wechsler Adult Intelligence Scale—Revised (WAIS-R)	1905-1970	26-87	94	1%	53	27%	117	47%	7	2%

TOTAL	94	100%	198	100%	248	100%	460	100%

Symbol Digit (Digit Symbol). Five studies (SATSA, GENDER, OCTO-Twin, LSADT, and MADT) administered a Symbol Digit task, while a sixth (MTSADA) administered a Digit Symbol task (total N = 8,757). Both tasks measure perceptual speed and accuracy, and so were pooled in analyses reported here. For convenience we designate this combined task as Symbol Digit. The same task was administered in the three Swedish and two Danish studies. Respondents were given a key containing nine separate two-dimensional geometric symbols that were assigned specific digits between 1 and 9. They were then presented with rows of symbols and were asked to state out loud what the correct digit was for each symbol. This Symbol Digit task was administered in two blocks of 50 with 45 seconds allowed for each block and the score being the number of correct trials across the two blocks (varying from 0 to 100). In MTSADA, the Digit Symbol task from the WAIS-R (Weschler, 1981) was administered, so that the task was reversed from that in the other studies. That is, in MTSADA respondents were asked to write down the symbol that corresponded with each target digit. There was one single administration of 90-second duration with a total of 90 items possible.

Table S6: Symbol Digit test version by study across age groups

				Age Groups							
		Birth Year	Age	<50		50-59		60-69		70+	
Studies	Test	Range	Range	N	%	N	%	N	%	N	%
SWEDISH & DANISH	Wechsler Adult Intelligence Scale (WAIS)	1900-1952	45-93	328	74%	960	95%	740	82%	846	97%

MTSADA	Wechsler Adult Intelligence Scale—Revised (WAIS-R)	1904-1970	26-87	116	26%	55	5%	162	18%	28	3%
TOTAL				444	100%	1015	100%	902	100%	874	100%

Vocabulary. Two studies (VETSA and MTSADA) administered a Vocabulary test in which respondents were required to generate definitions of words (total N = 2,030). In VETSA, the Vocabulary subtest from the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999) was administered, while MTSADA administered the Vocabulary subtest from the WAIS-R (Weschler, 1981).

Table S7: Vocabulary test version by study across age groups

				Age Groups							
		Birth Year	Age	<50		50-59		60-69		70+	
Studies	Test	Range	Range	N	%	N	%	N	%	N	%
VETSA	Wechsler Adult Intelligence Scale (WAIS)	1944-1955	51-59	0	0%	605	92%	0	0%	0	0%
MTSADA	Wechsler Adult Intelligence Scale—Revised (WAIS-R)	1905-1970	26-87	117	100%	56	8%	163	100%	27	100%
TOTAL				117	100%	661	100%	163	100%	27	100%

Synonyms. Five studies (SATSA, GENDER, OCTO, TOSS and VETSA) administered a Synonyms test (total N = 4,525). In all cases, the Synonyms test required the respondent to select the word that provided the best synonym to a target word from a set of alternatives. The specific words, alternatives and number of items varied across studies.

Table S8: Synonyms test version by study across age groups

		Age Groups			
Birth Year	Age	<50	50-59	60-69	70+

Studies	Test	Range	Range	N	%	N	%	N	%	N	%
SWEDISH	Wechsler Adult Intelligence Scale (WAIS)	1900-1948	50-91	0	0%	138	15%	133	100%	389	100%
TOSS	TOSS- Synonym subscale	1943-1971	32-59	702	100%	155	17%	0	0%	0	0%
VETSA	Armed Forces Qualification Test (AFQT)	1944-1955	51-59	0	0%	610	68%	0	0%	0	0%
TOTAL				702	100%	903	100%	133	100%	389	100%

IV. Calculating Observed Power

All model fits were compared to the no-moderation model to determine improvement in fit and a change in χ^2 was the difference by the estimated -2lnl for each model fit. Each model fit was assumed to represent the population estimates. Therefore, when compared to no-moderation model, the change in χ^2 could be used to find the non-centrality parameter (NCP) if certain moderation parameters were dropped. Before the χ^2 could be used to estimate power, the observed χ^2 needed to be adjusted by subtracting the dfs to determine the methods-of-moments estimator of the NCP. Since the observed χ^2 difference is a function by sample size, the ratio NCP by sample (NCP/N) was estimated by dividing the adjusted χ^2 difference by the total number of intact pairs in the analysis. The median NCP was then found

across all hypothesis tests, except for Digit Symbol, as the models tested were different (ADE), and Vocabulary, as this was only test to demonstrate no-moderation model to be the best fitting. The normed NCP was then used to find how much power was achieved for each cognitive test. A vector of NCP values was then generated by multiplying NCP/N by a series of trial samples sizes (i.e., NCP/N * N) and plotted on the inverse chi-squared distribution for the specific DF test with an alpha level of .05. Since plotting the iterated NCP demonstrated the full range of power on the NCP scale, each test was able to find the sample size required to achieve adequate power at 80%. Using a similar method, observed power was determined by the amount of power achieved on this distribution for total sample size within each cognitive test.

Table S9. Fit statistics for age-moderation models

Cognitive Test	Model Fit				Improvement in Fit Relative to No-Age- Moderation Model			Observed NCP/N	M te
	-2lnl	df	K	AIC	χ^2	Δdf	p		
Digits Forward									
1.No Age Moderation	74071.0	10084	13	74097.0					
2. Scalar model	74054.5	10083	14	74082.5	16.5	1	<.001	0.0024	
3. Linear Moderation only on A	74050.1	10083	14	74078.1	20.9	1	<.001	0.0031	
4. Linear Moderation only on C	74054.3	10083	14	74082.3	16.7	1	<.001	0.0025	
5. Linear Moderation only on E	74060.5	10083	14	74088.5	10.5	1	<.001	0.0015	
6. Linear Moderation on ACE	74050.1	10081	16	74082.1	20.9	3	<.001	0.0028	
7. Full ACE Moderation	74047.9	10078	19	74085.9	23.1	6	<.001	0.0027	
Digits Backward									
1.No Age Moderation	83171.5	11297	15	83201.5					
2. Scalar model	83159.8	11296	16	83191.8	11.7	1	<.001	0.0025	
3. Linear Moderation only on A	83155.9	11296	16	83187.9	15.6	1	<.001	0.0035	
4. Linear Moderation only on C	83158.3	11296	16	83190.3	13.2	1	<.001	0.0029	
5. Linear Moderation only on E	83163.6	11296	16	83195.6	7.9	1	0.004	0.0016	
6. Linear Moderation on ACE	83155.9	11294	18	83191.9	15.6	3	<.001	0.0030	
7. Full ACE Moderation	83153.5	11291	21	83195.5	18	6	0.006	0.0028	
Block Design									
1.No Age Moderation	17113.9	2306	10	17133.9					
2. Scalar model	17113.2	2305	11	17135.2	0.6	1	0.44		
3. Linear Moderation only on A	17106.0	2305	11	17128	7.8	1	0.005	0.0068	
4. Linear Moderation only on C	17099.7	2305	11	17121.7	14.2	1	<.001	0.0132	
5. Linear Moderation only on E	17109.8	2305	11	17131.8	4	1	0.05	0.0030	
6. Linear Moderation on ACE	17087.4	2303	13	17113.4	26.4	3	<.001	0.0234	
7. Full ACE Moderation	17087.0	2300	16	17119	26.8	6	<.001	0.0208	
Symbol Digit									
1.No Age Moderation	64409.3	8711	14	64437.3					
2. Scalar model	64409.0	8710	15	64439	0.3	1	0.58		
3. Linear Moderation only on A	64408.1	8710	15	64438.1	1.17	1	0.23	0.0000	
4. Linear Moderation only on D	64405.7	8710	15	64435.7	3.61	1	0.06	0.0006	
5. Linear Moderation only on E	64409.2	8710	15	64439.2	0.04	1	0.84		
6. Linear Moderation on ADE	64404.3	8708	17	64438.3	5	3	0.17	0.0004	
7. Full ADE Moderation	64381.8	8705	20	64421.8	27.4	6	<.001	0.0048	
8. Linear and Quadratic Moderation only on A	64404.7	8709	16	64436.7	4.5	2	0.03	0.0006	
9. Linear and Quadratic Moderation only on D	64386.5	8709	16	64418.5	22.8	2	<.001	0.0046	
10. Linear and Quadratic Moderation only on E	64409.2	8709	16	64441.2	0.1	2	0.98		

Vocabulary								
1.No Age Moderation	14613.2	2023	7	14627.2				
2. Scalar model	14613.1	2022	8	14629.1	0.1	1	0.78	
3. Linear Moderation only on A	14613.1	2022	8	14629.1	0.1	1	0.73	
4. Linear Moderation only on C	14613.1	2022	8	14629.1	0.1	1	0.75	
5. Linear Moderation only on E	14613.1	2022	8	14629.1	0.1	1	0.73	
6. Linear Moderation on ACE	14610.5	2020	10	14630.5	2.7	3	0.44	
7. Full ACE Moderation	14604.0	2017	13	14630	9.2	6	0.16	0.0033
Synonyms								
1.No Age Moderation	33136.5	4503	12	33160.5				
2. Scalar model	33021.2	4502	13	33047.2	115.3	1	<.001	0.0538
3. Linear Moderation only on A	33030.5	4502	13	33056.5	106	1	<.001	0.0494
4. Linear Moderation only on C	33065.6	4502	13	33091.6	70.9	1	<.001	0.0329
5. Linear Moderation only on E	33030.6	4502	13	33056.6	105.9	1	<.001	0.0494
6. Linear Moderation on ACE	33004.4	4500	15	33034.4	132.1	3	<.001	0.0608
7. Full ACE Moderation	32999.4	4497	18	33035.4	137.1	6	<.001	0.0617

k is the number of estimated parameters, including means.
AIC is equal to -2lnl+2*k
df = degrees of freedom
NCP/N=Non-centrality parameter divided by intact pairs
A = additive genetic, C = shared environmental, E = non-shared environmental, D = dominance. ACE models fit for all tests except Symbol Digit for which preliminary data suggested non-additivity so an ADE model was fit

1) a no-age-moderation model, 2) a scalar model is only moderation on the total phenotypic variance and all component are kept constant, 3-5) a linear age-moderation model only on one component (i.e., A, C or D, E and no quadratic terms), 6) a linear age-moderation model on the ACE or ADE, 7) a full age-moderation model, including both linear and quadratic term